区分的3次関数を用いたデータ平滑化 : 節点の決定について (計算の手間とデータ構造)

吉本, 富士市; 市田, 浩三; 清野, 武

数理解析研究所講究録
250巻 1975年 51-61

区分的3次関数を用いたデータ平滑化
―― 節点の決定について ――

明石高専　吉本富士市
京大　工　市田　浩三
京大　工　清野　武

1．まえがき

区分的な関数（スプライン関数）は，複雑な関数形をしたデータとか連続関数の近似関数としてよく用いられている．しかし，よい近似を得るためには節点を適切に決める必要がある[4)]．これは重要な問題の１つであるが，いまだ十分には検討されていないように思われる．ここでは，区分的3次関数を用いて最小２乗法によりデータを平滑化する場合の節点の決定について述べる．データ（の意味する関数）の２次微分を推定し，それを用いて節点の位置を求める．節点の数の決定は，誤差分散の不偏推定量を利用して行う．

2．節点のよい位置

いくつかの節点が与えられたとき，それをよい近似が得られる位置へ持ってくるためには，節点を変数にして問題を非

線形最小2乗法にして解くことが考えられる[2]. しかし，これは極値が多くあるので，節点のよい初期推定値を選ばないと計算量が多く，またよい近似にならない．ところで，1次微分まで連続な（C^1）区分的3次関数は，2次微分がつなぎ目で不連続な折れ線になる特徴がある．また，あるデータをつなぎ目で不連続な折れ線で近似することは比較的容易である．近似関数として1次微分まで連続な区分的3次関数を用いたとき，よい近似を得るためにはデータ（の意味する関数）の2次微分と近似関数の2次微分ができるだけ近くなるように考慮すればよいと考えられる．すなわち，簡単に考えると次のようになる．

区間$[x^{(i-1)}, x^{(i)}]$，$(i=1, 2, \cdots, n)$に対して区分的3次関数を次式で表わす[1].

$$S_i(x) = m_{i-1}\frac{(x^{(i)}-x)^2(x-x^{(i-1)})}{h_i^2} - m_i\frac{(x-x^{(i-1)})^2(x^{(i)}-x)}{h_i^2}$$

$$+y_{i-1}\frac{(x^{(i)}-x)^2\{2(x-x^{(i-1)})+h_i\}}{h_i^3} + y_i\frac{(x-x^{(i-1)})^2\{2(x^{(i)}-x)+h_i\}}{h_i^3}$$

$$\equiv A_i x^3 + B_i x^2 + C_i x + D_i \,. \tag{1}$$

ここに，$h_i = x^{(i)} - x^{(i-1)}$であり，$y_i$，$m_i$はそれぞれ節点$x^{(i)}$での関数値と微分の値である．すると

$$\frac{dS_i(x)}{dx} = m_{i-1}\frac{(x^{(i)}-x)(2x^{(i-1)}+x^{(i)}-3x)}{h_i^2} - m_i\frac{(x-x^{(i-1)})(2x^{(i)}+x^{(i-1)}-3x)}{h_i^2}$$

$$+\frac{y_i - y_{i-1}}{h_i^3}\,6\,(x^{(i)}-x)(x-x^{(i-1)})$$

$$= 3A_i x^2 + 2B_i x + C_i \tag{2}$$

$$\frac{d^2S_i(x)}{dx^2} = -2m_{i-1}\frac{2x^{(i)}+x^{(i-1)}-3x}{h_i^2} - 2m_i\frac{2x^{(i-1)}+x^{(i)}-3x}{h_i^2}$$

$$+6\,\frac{y_i - y_{i-1}}{h_i^3}(x^{(i)}+x^{(i-1)}-2x)$$

$$= 6A_i x + 2B_i \tag{3}$$

となる．いま，ある関数$f(x)$を(1)で近似するとしよう．このとき，$f''(x) \simeq S_i''(x)$とすれば，C_iをうまく決めることにより$f'(x) \simeq S_i'(x)$とできる．反対に，$f''(x)$と$S_i''(x)$がかなり違っていると$f'(x) \simeq S_i'(x)$とできない．なぜなら，(2)よりわかるように$S_i'(x)$はC_iを変化するだけでは，その‘形’はかわらず‘位置’が変化するだけであるからである．同様に$f'(x) \simeq S_i'(x)$とできれば，D_iをうまく決めることにより$f(x) \simeq S_i(x)$とできる．そこで，データ（の意味する関数）の2次微分を推定し，それをつなぎ目で不連続な折れ線

で近似すれば，そのつなぎ目が節点のよい位置になると思われる．

3．2次微分の推定

一般に，誤差のあるデータ（の意味する関数）の2次微分を精度よく求めることは容易でない．ここでは，最小2乗法を用いた移動平均法を利用してそれを求める．このときの近似関数には3次関数を用いて，その2次微分を求め2次微分の推定値とする．各データ点での2次微分を計算する．最小2乗法は，各点において，その両側の同じ個数のデータを考慮する．データが(x_k, f_k)，$(k=1, 2, \cdots, N)$と与えられており，f_kにのみ誤差があると仮定する．いま，データ点x_jにおける2次微分を推定しよう（図1）．近似関数を

$$P_3(x)=a_0+a_1x+a_2x^2+a_3x^3 \qquad (4)$$

とする．データf_jの両側の考慮するデータの個数をα_j個としたとき，残差の2乗和は

$$R=\sum_{k=j-\alpha_j}^{j+\alpha_j}\left\{P_3(x_k)-f_k\right\}^2 \qquad (5)$$

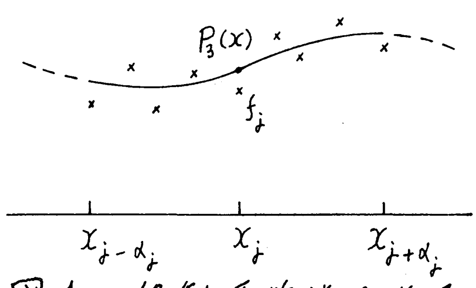

図1　移動平均法による2次微分の推定

となる．これを最小にするパラメータa_2，a_3を求めると

$$\left.\frac{d^2 P_3(x)}{dx^2}\right|_{x=x_j} = 2a_2 + 6a_3 x_j \tag{6}$$

となる．移動平均法では，普通 α_j をある値に固定している．しかし，適切な α_j の値はデータによって異なるので，よい α_j を求めることは簡単でない．そこで，‘傾向(trend)’[3] を用いて α_j を各データ点においてよい値に自動的に決める．残差を

$$Z_k = P_3(x_k) - f_k \tag{7}$$

とするとき，傾向の判定条件は

$$\sum_{k=j-\alpha_j+1}^{j+\alpha_j} Z_k Z_{k-1} \geq \beta R \sqrt{2\alpha_j} \Big/ (2\alpha_j + 1) \tag{8}$$

となる．この式の右辺は左辺の標準偏差をβ倍したものである．各点 x_j において，(8) が満足されない最も大きな α_j に対する (6) の値を，その点の2次微分とする．始めと終りのいくつかのデータについては，このような扱いができないので特別に扱う．βの値は，数値実験によれば0.5程度が望ましい．なお，データが等間隔であれば原点移動を行うことにより計算量を少なくできる．

4．2次微分の近似

推定した2次微分を，つなぎ目で不連続な折れ線で近似す

る．データ点 x_k に対する推定した2次微分を f''_k と書く（$k=1, 2, \cdots, N$）．許容誤差 ε_{max} を与えて，f''_k に対する最小2乗近似を求める．まず，最初の3個の2次微分値に線分 $a_1 x + b_1$ をあてはめて，残差2乗和

$$U_1 = \sum_{k=1}^{3} \left\{ a_1 x_k + b_1 - f''_k \right\}^2 \tag{9}$$

を計算する．もしも $U_1 > \varepsilon_{max}$ であれば，最初の線分は点（x_1，f''_1）と（x_2，f''_2）を通る直線である．そうでなければ（x_4, f''_4）を含めて再度最小2乗近似を行う．このとき，前の結果が使えるので計算量は少ない．このようにして，考慮する2次微分値を1つずつ増大しながら $U_1 > \varepsilon_{max}$ となる点（x_{J+1}，f''_{J+1}）をさがすと，最初の線分は（x_k，f''_k），（$k=1$，2，…，J）に対する最小2乗近似線分となる．次に，（x_{J+1}，f''_{J+1}）から始めて2番目の近似線分 $a_2 x + b_2$ を計算する．このプロセスを，すべての2次微分値に対する近似線分が求まるまでくり返すと，2次微分は何本かの線分で近似され，その各線分の残差2乗和は ε_{max} 以下である．

以上の方法によって得られた結果は最適なものではない．そこで，線分の本数を固定して，各区間の残差2乗和のうち最大なものをできるだけ小さくするような最適化を行う．

近似線分の本数は，許容誤差を変化すれば変えられる．ま

ず，許容誤差を

$$\varepsilon_{max} \geq N\left[\left\{\max_{1\le k\le N}\left(f''_k\right) - \min_{1\le k\le N}\left(f''_k\right)\right\}\Big/2\right]^2 \tag{10}$$

となる値にすれば，すべてのデータに対して線分数1の最小2乗近似を得る．次に最適化後の残差2乗和より少し小さい値を許容誤差にすれば，線分数2の近似となる．以後，最適化後の最大の残差2乗和より少し小さい値を次の許容誤差にするプロセスを続ければ，各線分数の場合の最小2乗近似を得る．この線分のつなぎ目を節点の位置とすれば，各節点数の場合のよい節点の位置となると思われる．以上のようにすれば，各区間には2個以上のデータが含まれている．

5．区分的3次関数を用いた最小2乗法

データに対して(1)を最小2乗法によりあてはめる（図2)．残差の2乗和は

$$Q = \sum_{i=1}^{n} \sum_{k=p_i}^{q_i} \left\{ S_i(x_k) - f_k \right\}^2 \tag{11}$$

となる．(11)を(y_i，m_i)，($i=0, 1, \cdots, n$)で偏微分して0とおけば正規方程式

$$\mathbf{D}\mathbf{z} = \mathbf{g} \tag{12}$$

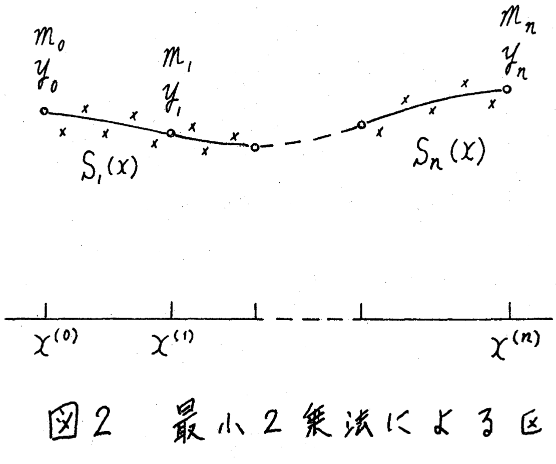

図2　最小2乗法による区分的3次関数のあてはめ

を得る．ここに，

$$\mathbf{Z} = (y_0, m_0, \cdots, y_n, m_n)^T \tag{13}$$

である．(12)の係数行列Dは対称な帯行列となる特長がある．(12)を解いて(1)へ代入すれば近似関数を求められる．よい節点の数は以下のようにして決める．(1)より，次の回帰模型を考える．

$$f_k = S_i(x_k) + \varepsilon_k \;,\; (x^{(i-1)} \leq x_k \leq x^{(i)},\; i = 1, 2, \cdots, n). \tag{14}$$

ここに，ε_k は平均値0，分散σ^2の誤差であるとする．最小2乗法により$S_i(x)$をデータへあてはめると，誤差分散の不偏推定量は次のようになる．

$$\delta = Q \big/ (N - 2n - 2). \tag{15}$$

4節で求めた節点の位置を用いてあてはめ計算すると，節点数の増加とともにδは小さくなっていき，あるところで落着くようになる（図5参照）．このとき，よいあてはめができており適切な節点数と考えられる．

4節で述べたように，各区間には2個以上のデータが含まれているので，(12)は常に一意的な解を持つようになる[5]．以上の方法によるデータ平滑化の結果が十分でない場合にも，

こうして得られた節点を初期値とすれば，よい初期値となっているので，簡単な最適化のアルゴリズムにより十分よい結果を得ることができる．

6．計算例

次のような関数形をしたデータを用いた．

$$f_k = \frac{1}{0.01 + 10(x_k - 0.3)^2} + \frac{1}{0.015 + 10(x_k - 1.2)^2} + \varepsilon_k . \qquad (16)$$

ここに，ε_k は正規分布をする誤差で，平均値 0，分散 1 である．データ点 x_k は 0.005 (0.01) 1.995 の 200 個とした．図3は推定した2次微分を折れ線により近似したものである．その折れ線のつなぎ目に節点を入れて最小2乗近似すると図4となる．図5は，ここで述べた方法の

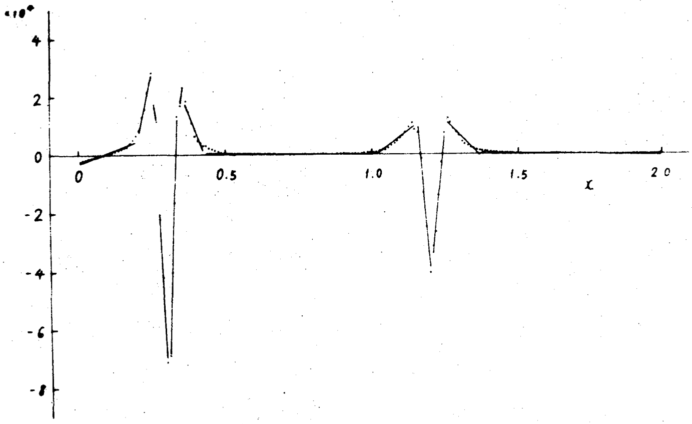

図3　2次微分の近似

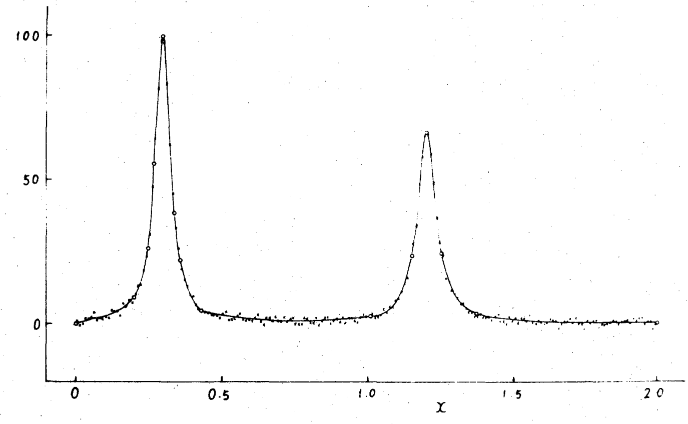

図4　2次微分を用いて決めた節点による計算結果（節点数14）

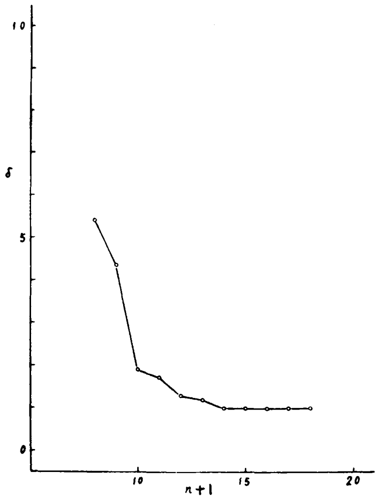

図5　誤差分散の不偏推定量（2次微分利用）

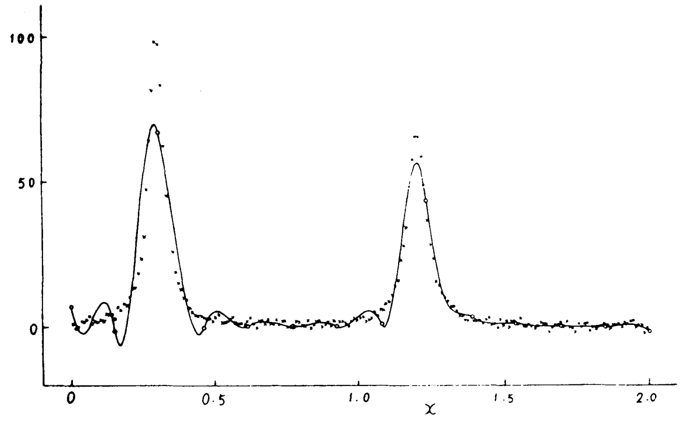

図6　等間隔節点による計算結果（図4と同じ節点数）

場合の誤差分散の不偏推定量を示す．図6は，図4と同じ節点数で等間隔に節点を入れた場合である．等間隔節点のときの誤差分散の不偏推定量は図7のようになる．ここで，図7の目盛は図5の10倍であることに注意したい．

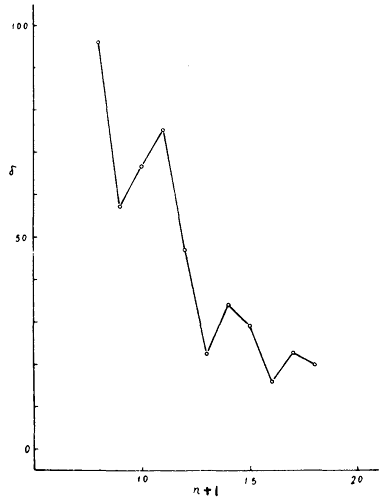

図7　誤差分散の不偏推定量（等間隔節点）

7．あとがき

以上において，節点の数と位置を適切に決めた区分的3次関数を用いてデータを平滑化する方法を述べた．区分的3次関数を2次微分まで連続にする（C^2）場合には，2次微分を折れ線近似するとき，つなぎ目を連続にして考える必要があ

る．これはかなりむずかしいので今後の問題としたい．何階かの微分を利用して節点を決める考え方は，他にも利用できると思われる．なお，3次微分または4次微分を利用して節点を決定することも考えられるが，一般に高階微分になるほど推定が困難であるので，ここでは検討しなかった．3節で述べた方法では，3次微分以上になると推定精度があまりよくなかった．

参考文献

1) J. H. Ahlberg, E. N. Nilson and J. L. Walsh: The Theory of Splines and Their Applications, P. 284, Academic Press, London (1967).

2) C. de Boor and J. R. Rice: Least squares cubic spline approximation II, Purdue University Reports, CSD TR 21 (1968).

3) M. J. D. Powell: Curve fitting by splines in one variable, Numerical approximation to functions and data, J. G. Hayes, ed., PP. 65-83, Athlone Press, London (1970).

4) J. R. Rice: The Approximation of Functions Vol. II, P. 334, Addison-Wesley, Reading, Massachusetts (1969).

5) I. J. Schoenberg and A. Whitney: On Pólya frequency functions III, Trans. Amer. Mathem. Soc., Vol. 74, PP. 246-259 (1953).